

# UCSF

## UC San Francisco Previously Published Works

### Title

Structural basis for cooperative RNA binding and export complex assembly by HIV Rev.

### Permalink

<https://escholarship.org/uc/item/8md5z8rt>

### Journal

Nature structural & molecular biology, 17(11)

### ISSN

1545-9993

### Authors

Daugherty, Matthew D

Liu, Bella

Frankel, Alan D

### Publication Date

2010-11-01

### DOI

10.1038/nsmb.1902

Peer reviewed



Published in final edited form as:

*Nat Struct Mol Biol.* 2010 November ; 17(11): 1337–1342. doi:10.1038/nsmb.1902.

## Structural basis for cooperative RNA binding and export complex assembly by HIV Rev

Matthew D. Daugherty<sup>1</sup>, Bella Liu<sup>2</sup>, and Alan D. Frankel<sup>2,\*</sup>

<sup>1</sup> Chemistry and Chemical Biology Graduate Program, University of California, San Francisco San Francisco, CA 94158

<sup>2</sup> Department of Biochemistry and Biophysics, University of California, San Francisco San Francisco, CA 94158

### Abstract

HIV replication requires nuclear export of unspliced viral RNAs to translate structural proteins and package genomic RNA. Export is mediated by cooperative binding of the Rev protein to the Rev response element (RRE) RNA, forming a highly specific oligomeric ribonucleoprotein (RNP) that binds to the Crm1 host export factor. To understand how protein oligomerization generates cooperativity and specificity for RRE binding, we solved the crystal structure of a Rev dimer at 2.5 Å resolution. The dimer arrangement organizes arginine-rich helices at the ends of a V-shaped assembly to bind adjacent RNA sites, structurally coupling dimerization and RNA recognition. A second protein–protein interface arranges higher-order Rev oligomers to act as an adapter to the host export machinery, with viral RNA bound to one face and Crm1 to another, thereby using small, interconnected modules to physically arrange the RNP for efficient export.

### Introduction

Retroviruses such as HIV are encoded in small RNA genomes that contain multiple splice sites. This poses a problem to virus replication because unspliced RNAs, including those that code for the viral structural proteins and genomic RNA, typically are retained in the nucleus<sup>1</sup>. To overcome the problem, HIV utilizes the Rev protein to bind to and oligomerize on the highly structured ~350 nt Rev response element (RRE) RNA located in viral introns, forming a large RNP that directs transport of fully and partially unspliced RNAs to the cytoplasm<sup>1,2</sup>. The assembled Rev–RRE ribonucleoprotein (RNP) binds to the host export factor Crm1 (Xpo1) and is shuttled through the nuclear pore complex, releasing the RRE

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Contact: Alan Frankel, Department of Biochemistry and Biophysics, UCSF, 600 16<sup>th</sup> St., San Francisco, CA 94158, 415-476-9994, frankel@cgl.ucsf.edu.

**Accession code:** Atomic coordinates and structure factors have been deposited with the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) under accession code 3LPH.

**Author Contributions:** M.D.D. and A.D.F. designed experiments and analyzed data. M.D.D. performed RNA-binding and SEC experiments. M.D.D. and B.L. purified and crystallized Rev and collected diffraction data. M.D.D. solved and refined the structure. M.D.D. and A.D.F. wrote the manuscript.

The authors declare no conflict of interest.

containing RNAs into the cytoplasm and allowing reimport of Rev into the nucleus for further rounds of nuclear export<sup>1–3</sup>.

Assembly of the Rev–RRE RNP, and thus progression through the viral life cycle, requires the cooperative addition of Rev monomers along the RRE in order to form a high-affinity complex<sup>4–6</sup>. The Rev homo-oligomer forms an exquisitely specific complex with the RRE by making multiple contacts between arginine rich motifs (ARMs) (Fig. 1A) of individual Rev monomers and several different binding sites within the RRE, resulting in a discrete, hexameric complex with 500-fold higher affinity than any monomeric Rev–RNA complex<sup>6,7</sup>. One well-characterized RNA site, known as stem IIB, is required for specific RRE recognition and initiates assembly of the oligomer<sup>8–13</sup>. This ARM–IIB interaction is understood at the molecular level from NMR studies<sup>14</sup>. Interestingly, the only other characterized binding site, known as stem IA, is recognized by a different surface of the  $\alpha$ -helical ARM, taking advantage of the inherent adaptability of this RNA-binding motif<sup>6</sup>. In fact, the RNA-binding ARM of Rev has been shown to bind to many different selected nucleic acid structures and sequences<sup>15–17</sup>, making it a mystery how such a promiscuous domain can be employed to recognize HIV RNA with such high specificity.

Protein–protein interactions between Rev monomers are crucial for Rev binding specificity and RNA export and are mediated by oligomerization domains that flank the ARM<sup>4</sup>. Indeed, a single point mutation in an oligomerization domain results in non-cooperative complex formation and lowers affinity by two orders of magnitude<sup>6</sup>, demonstrating the plausibility of Rev oligomerization as a target for HIV inhibitor design. On an extended IIB RNA, in which a helical disruption is oriented proximal to the primary IIB site, oligomerization mediates cooperative binding of a Rev dimer and increases specificity compared to the monomer–IIB interaction<sup>6,18</sup>. Detailed biochemical studies revealed that oligomerization relies on two distinct sets of hydrophobic amino acids; one set, including Leu18 and Ile55, is required to form a cooperative Rev–RNA dimer, and the other set, including Leu12 and Leu60, to form higher-order complexes<sup>19</sup>. A recent crystal structure of Rev bound to a monoclonal Fab fragment revealed the details of the higher-order oligomerization surface utilizing Leu12 and Leu60, but the cooperative dimerization surface was blocked by Fab binding<sup>20</sup>.

To understand how the organization of monomers within the Rev oligomer facilitates highly specific recognition of HIV RNA and export complex assembly, we solved the crystal structure of a Rev dimer that cooperatively binds the RRE, and additionally observed higher-order oligomerization of dimers in the crystal. This structure allows us to complete the model of the Rev oligomer, showing how protein oligomerization is physically coupled to cooperative RNA recognition. Burial of the hydrophobic oligomerization domains upon dimerization determines the orientation of extended ARMs toward RNA, imparting Rev with the spatial specificity needed to recognize the RRE and augment the broad specificity of the individual ARMs. Higher-order oligomerization of Rev dimers also serves to create a nuclear export adaptor, where the viral RNA binds to one side of the oligomer and Crm1 to the other.

## Results

### Structure determination and overview

To obtain high-quality crystals of a Rev dimer capable of binding cooperatively to RNA, we partially disrupted the higher-order oligomerization surface that otherwise leads to protein aggregation by mutating positions 12 and 60, and also removed 46 residues from the disordered C-terminus<sup>7,21</sup> (Fig. 1a). The resulting protein, termed Rev<sub>70</sub>-Dimer, binds to a stem IIB RNA long enough to accommodate two Rev monomers with cooperativity and high affinity (Fig. 1b). The protein retains a strong oligomerization interface, behaving as a dimer even in the absence of RNA (Fig. 1c).

The Rev<sub>70</sub>-Dimer structure was solved using multi-wavelength anomalous diffraction (MAD) to a final resolution of 2.5 Å (Table 1 and Supplementary Fig. 1). The asymmetric unit contains two copies of the Rev dimer arranged front to back (Fig. 2a). Each monomer adopts an antiparallel helix-loop-helix structure, as anticipated by biochemical and recent crystallographic studies<sup>19–21</sup>, with a nearly parallel orientation of helices that comprise the oligomerization domains (Figs. 2a–b). All four monomers in the asymmetric unit are nearly superimposable, with a well-folded core domain (residues 9–63) in all cases (Fig. 2b). The two monomers in each dimer are arranged face to face in a V-shaped topology, with the base of the V forming a central interface comprised of four helical oligomerization domains that buries residues known to be essential for dimerization<sup>19</sup>, and the RNA-binding ARMs extending out at an angle of ~120° relative to each other (Fig. 2a).

### Conservation of the Rev monomer fold

The Rev monomer is stabilized by a conserved network of hydrophobic and polar residues that form intramolecular contacts across the two oligomerization domains. Four branched nonpolar residues (Ile19, Leu22, Ile52, and Ile59) form the hydrophobic core of the monomer (Fig. 2a) and are crucial for RRE binding<sup>19</sup>. Each monomer contains a hydrogen bonding network in which the sidechain of Asn26 is held between the guanidinium group of Arg48 and the carbonyl of the Gln49 sidechain, precisely aligning the C-terminal end of helix 1 (Asn26) with the C-terminal end of the ARM in the middle of helix 2 (Figs. 2b–c). Additional hydrogen bonds between the side chains of Tyr23 and His53 and between the indole nitrogen of Trp45 and the backbone carbonyl of Asn26 stabilize this helical packing and help constrain the proline-rich loop connecting helix 1 and helix 2 (Fig. 2c). Residues that participate in this network of interactions are highly conserved among all HIV-1 isolates (Fig. 1a and 2b), indicating the importance of maintaining the oligomerization domains in proper register. Interestingly, Trp45, Arg48, and Gln49 are not strictly part of the oligomerization domains but rather reside in the ARM, and mutation of these residues disrupts assembly of higher order Rev RRE complexes<sup>6,19</sup>, underscoring the functional, as well as physical, coupling of oligomerization and RNA binding.

### Dimerization is physically coupled to binding cooperativity

The Rev dimer is formed by the packing of hydrophobic residues between monomers, creating an extensive interface that buries over 1500 Å<sup>2</sup> of the solvent accessible area of the dimer and totaling nearly 20% of the total protein surface area (Figs. 3a–b). Leu18 and

Ile55, which form symmetric contacts between monomers, and Phe21, which packs into a groove between the helices of the opposing monomer (Fig. 3c), are key residues at the dimerization interface and are highly conserved and essential for cooperative RNA binding and export<sup>6,19,22</sup>. The conserved structural and functional roles of these residues, as well as kinetic studies suggesting the dimer is an early intermediate in RNP assembly<sup>13</sup>, make the dimerization surface a plausible target for inhibitor design. It is especially interesting that hydrophobic residues important in the monomer structure, particularly Leu22 and Ile59, simultaneously contribute to the dimer interface by being buried in the groove formed between the helices of the partner monomer (Fig. 3d). In the absence of dimerization, these residues would be nearly 50% solvent exposed, thus tightly linking the hydrophobic interfaces in the monomer and dimer. The burial of hydrophobic surfaces of the monomers is estimated to stabilize the dimer by >22 kcal/mol<sup>23</sup>, providing one structural explanation for cooperative RNA binding and suggesting that a fundamental building block of the Rev RRE complex is a protein dimer.

An even more important structural explanation for cooperative RNA binding arises from the orientation of the ARMs specified by the dimer interface. Previous studies have shown that an extended stem IIB RNA can cooperatively bind two Rev monomers only if protein oligomerization is intact and the RNA sites are properly positioned relative to each other<sup>6,18</sup>. The dimer structure provides a structural rationale for these results, revealing that the quaternary fold defines the angle and position of the RNA-binding surfaces (Fig. 3e), with the two ARMs arranged to reach out from the body of the dimer and grasp RNA, much like two human arms are positioned to grip objects. Such an orientation of ARMs would specify the organization of RNA sites required for cooperative complex formation. Indeed, placing the IIB RNA hairpin from the NMR structure of the ARM peptide-RNA complex<sup>14</sup> (labeled IIB34) into the crystal structure, and extending the RNA by four additional base pairs to accommodate a second Rev monomer<sup>6</sup>, demonstrates how bipartite RNA binding would stabilize Rev dimerization (Fig. 3e). The arrangement of this complex is consistent with the molecular envelope of a similar Rev dimer-RNA complex determined by small angle X-ray scattering (SAXS)<sup>7</sup>. Most interestingly, we also can infer that different faces of the helical ARM are positioned to contact the RNA from each monomer, consistent with the observation that different amino acids are used to recognize stem IIB and stem IA sites<sup>6</sup>. In our model, the first ARM utilizes Asn40 (yellow) on the ‘inner face’ of the helix to contact IIB<sup>14</sup>, leaving the end of IIB available for binding by the ‘outer face’ of the second ARM (Fig. 3e) and explaining how oligomerization pairs the binding adaptability of the ARM to the spatial organization of binding sites in the RRE. This arrangement leads to cooperative RNP formation and provides the structural rationale for how Rev distinguishes RRE-containing viral RNAs from host RNAs using the promiscuous ARM RNA-binding motif.

### A model for Rev RRE export complex assembly

The crystal structure also indicates how higher order complexes assemble from Rev dimer building blocks. The protein that crystallized contains two mutations (L12S and L60R) meant to weaken the higher-order oligomerization interface<sup>19</sup>, and indeed behaves biochemically as a dimer (Fig. 1). However, in the asymmetric unit and throughout the crystal lattice (Supplementary Figs. 2–4), we observe packing of dimers using the expected

higher-order surface containing residues 12, 16, and 60<sup>19</sup>. The higher-order arrangement of Rev dimers observed between asymmetric units in our crystal is nearly identical to that recently observed in a monoclonal Fab-Rev complex (Figs. 4a and Supplementary Fig. 4), where the Fab blocks the dimerization surface<sup>20</sup>. These data, and our recent finding that six Rev monomers assemble on the RRE<sup>7</sup>, permit us to model a Rev hexamer (Fig. 4b) and to propose an arrangement for the export-competent RNP (Fig. 4c). The disordered nuclear export sequences (NESs; residues 73–83)<sup>7,21</sup> that interact with Crm1 are not present in our crystal structure but originate at the end of helix 2 and thus are predicted to project away from the ARMs of the Rev hexamer, like six tentacles of a jellyfish. The arrangement of this Rev–RRE RNP positions the NESs for Crm1 binding but is unlikely to accommodate more than one or two Crm1–RanGTP complexes (Fig. 4c), consistent with the hypothesis that Rev oligomerization is needed to enhance RNA-binding affinity and not necessarily to recruit additional Crm1 complexes<sup>6</sup>. This model suggests that Rev functions as a molecular adapter, with RNA bound to one face of the oligomer and Crm1 to the opposite face, providing a simple architecture to facilitate interactions with the nuclear pore and promote RNA export.

## Discussion

The switch from early to late stages of HIV replication is facilitated by Rev-mediated export of unspliced viral RNAs. Rev assembles on the RRE with surprising complexity, utilizing a promiscuous RNA-binding ARM to recognize multiple different RNA sites cooperatively in a process that is dependent on protein oligomerization<sup>6</sup>. The high-resolution Rev oligomer structure presented in this work reveals the physical details to explain how oligomerization mediates specific, cooperative RRE binding and export complex formation. The functional link between cooperativity and RNA export can be viewed as a direct product of the physical coupling between oligomerization and RNA binding (Fig. 5). RNP assembly is initiated by binding of a Rev monomer to the stem IIB site, employing the RNA-binding ARM to recognize the nucleotide sequence and backbone structure of the IIB element<sup>14</sup> (Fig. 5, left). This single interaction alone, however, is not sufficient to mediate export, and further specificity conferred by cooperative oligomerization is required for Rev function<sup>4–6</sup>. Cooperativity and an additional level of spatial specificity are added to RNP assembly by utilizing a composite binding site for the second Rev monomer which is formed by a secondary RNA site and the exposed dimerization interface (Fig. 5, middle). The structure reveals that cooperative assembly of the Rev dimer will proceed only if the spacing and orientation of the ARMs match the RNA binding sites, providing a rationale for previous biochemical results<sup>6,18</sup>, and reminiscent of the intricate interdependencies of cooperative ribosomal assembly<sup>24</sup>. Such an organization allows Rev to ‘measure’ the distance between RNA sites, thus augmenting the promiscuous binding specificity that has been described previously for the arginine-rich RNA-binding domain<sup>15–17</sup>. Further cooperativity and affinity are gained as the size of the RNP grows<sup>6</sup>, mediated by the higher order oligomerization interface and additional ARM-RNA contacts and allowing Rev to increasingly distinguish the RRE from other RNAs (Fig. 5, bottom). Given that the RRE specifies the size of the Rev oligomer in the RNP<sup>7</sup>, it will be interesting to determine at a structural level if RNA caps the ends of the Rev hexamer to block further subunit addition or affects the packing arrangement between Rev subunits. The discrete architecture of the

assembled hexameric Rev–RRE RNP, with its exposed NESs on one face of the oligomer and the viral RNA tethered to the other, is thus organized for efficient nuclear export.

The use of a promiscuous RNA-binding motif in the context of a homo-oligomer may facilitate Rev–RRE evolution in a rapidly changing virus such as HIV. The specificity achieved by orientating RNA sites within the RRE may allow more relaxed specificity at the level of the individual ARM–RNA interactions, providing a means both to evolve and escape functionally deleterious mutations within a given binding site. The framework of the oligomerization domains may likewise play a role; in some ways, the fold of the Rev oligomer resembles DNA-binding transcription factors that organize pairs of recognition helices to assemble on adjoining DNA sites<sup>25–27</sup>, but the Rev oligomerization interfaces utilize a rather obtuse helical crossing angle, reminiscent of  $\alpha$ -helical packing in globin proteins, to set the geometry of the two recognition helices. Analyses of the globin fold have suggested that such crossing angles might be primed to adapt during evolution, with substitutions allowed as long as they preserve the packing between the hydrophobic ridges of one helix and the grooves of the other helices<sup>28</sup>. The flexibility to evolve Rev oligomerization interfaces may couple to the adaptability of the RNA-binding ARMs themselves, allowing Rev to simultaneously optimize contacts between each ARM and its individual RRE-binding site, such as those in stems IIB and IA<sup>6,14</sup>, and the spatial arrangements between sites. Thus, the numerous adaptable interaction surfaces of the Rev–RRE complex both provide a means to evolve and retain function in a rapidly evolving virus and enable HIV to generate an exquisitely specific, cooperative, and robust RNP export complex using small, interconnected sequence motifs.

## Methods

### Protein expression and purification

To facilitate crystallization, the disordered C-terminal 46 residues of Rev were removed<sup>7,21</sup> and oligomerization domain mutations, L12S and L60R, were introduced to partially disrupt higher-order oligomerization<sup>19</sup>. The resulting Rev<sub>70-Dimer</sub> protein was expressed in *E. coli* strain BL21/DE3 from a pHGB1-derived vector with an N-terminal His<sub>6</sub> tagged GB1 domain followed by a TEV protease cleavage site<sup>6</sup>. Cells were grown in SBMX, a phosphate-based medium (described in Supplementary Methods) and expression was induced with 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside for 4 h at 37°C.

Rev was purified using Ni-NTA affinity chromatography as previously described<sup>7</sup> and dialyzed overnight at 4°C against crystallization buffer [40 mM Tris pH 8.0, 200 mM NaCl, 100 mM Na<sub>2</sub>SO<sub>4</sub>, 400 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>] containing 2 mM  $\beta$ -mercaptoethanol. To remove the GB1 domain, TEV protease was added and incubated at room temperature for 12 h and the reaction was loaded onto Ni-NTA resin equilibrated with crystallization buffer containing 20 mM imidazole. Rev<sub>70-Dimer</sub> was collected in the flow through and concentrated to 14 mg ml<sup>-1</sup> in crystallization buffer containing 2 mM dithiothreitol.



## RNA binding gel shift assays

*In vitro* transcription, radiolabeling, and electrophoretic mobility shift assays with IIB42 (5'-GGAGGUAUAUGGGCGCAGCGCAAGCUGACGGUACAGGCCUCC-3') were performed as described<sup>6</sup>. To prevent aggregation in low sulfate conditions, Rev<sub>70</sub>-Dimer was supplemented with equimolar amounts of yeast tRNA before serial dilution in buffer (5 mM Tris pH 8.0, 20 mM NaCl, and 200  $\mu\text{g ml}^{-1}$  bovine serum albumin). Binding constants were calculated by measuring the fraction of all bound RNA species compared to total RNA, and fitting the data to binding curves using Kaleidagraph software (Synergy Software, Reading, PA) using three replicates.

## SEC and MALS

SEC/MALS data were collected using an Ettan LC system (GE Life Sciences) with a silica gel KW803 column (Shodex) equilibrated in crystallization buffer at a flow rate of 0.35  $\text{ml min}^{-1}$ . The system was coupled on-line to an 18-angle MALS detector and a differential refractometer (DAWN HELEOS II and Optilab rEX, Wyatt Technology). Molar mass determination was calculated with ASTRA 5.3.1.5 software.

## Crystallization

Initial crystallization attempts with the dimeric Rev-IIB complex (Fig. 1b) yielded promising hits in ammonium sulfate conditions. It was subsequently determined that the crystals contained only protein, consistent with previous observations that ammonium sulfate stabilizes the protein in the absence of RNA<sup>7</sup> and thereafter the protein was purified in the presence of ammonium sulfate. Crystals were obtained by hanging drop vapor diffusion by mixing 1  $\mu\text{l}$  protein at 10–14  $\text{mg ml}^{-1}$  in crystallization buffer and 1  $\mu\text{l}$  reservoir solution containing 100 mM Tris pH 8.0, 50–100 mM NaCl, 1.45–1.55 M  $(\text{NH}_4)_2\text{SO}_4$ , and 3% (w/v) PEG 1000. After 10–14 days, cryoprotectant containing 1.8 M sodium malonate, 100 mM Tris pH 8.0, 500 mM NaBr and 250 mM  $\text{Na}_2\text{SO}_4$  was added incrementally and crystals were transferred to a drop containing cryoprotectant only, allowed to equilibrate for 15–60 minutes and flash frozen in liquid nitrogen. Crystals used for MAD phasing were treated identically utilizing a cryoprotectant containing  $\text{Na}_2\text{SeO}_4$  instead of  $\text{Na}_2\text{SO}_4$ .

## Data collection and structure determination

Data sets were collected at the Advanced Light Source beamline 8.3.1 on crystals maintained at 100K. The sodium malonate cryoprotectant, together with NaBr, reduced the c-axis of the unit cell from as long as 330 Å down to 81 Å and increased the symmetry of the crystals (Daugherty and Frankel, unpublished). Such changes, while unexpected, may be explained by the contacts within and between asymmetric units that appear to be mediated by  $\text{Br}^-$  and malonate ions. Diffraction data were processed and scaled with the HKL-2000 package<sup>32</sup>. Using MAD data on  $\text{SeO}_4$ -soaked crystals, Se atoms were located using SOLVE (see Supplementary Fig. 1) and initial models were built with RESOLVE<sup>33</sup> within PHENIX<sup>34</sup>.

Initially, loose non-crystallographic symmetry (NCS) restraints were applied to build the four monomers in the asymmetric unit using REFMAC5<sup>35</sup>. Subsequent iterative model



building and refinement were carried out using COOT<sup>36</sup> and PHENIX<sup>34</sup> without NCS restraints. The refined structure from the experimentally phased SeO<sub>4</sub> data was used as a molecular replacement model for the native data set using PHASER<sup>37</sup>. Thirteen Br atoms were located with PHASER using molecular replacement and SAD data from the Br edge. Excellent electron density can be seen for residues 9–63 in all four monomers, with variable ends to the electron density of each chain. Sidechains for which electron density was weak or lacking were not modeled. Data collection and refinement statistics are summarized in Table 1. All monomers had good stereochemistry and showed no Ramachandran outliers<sup>38</sup>. Solvent accessible areas were calculated using AREAIMOL in the CCP4 suite<sup>39</sup>. Graphic figures were generated using PyMOL<sup>40</sup>.

A structural model of the Rev dimer with stem IIB was generated using the NMR structure of a stem IIB–ARM peptide complex<sup>14</sup> (PDB code 1ETF). The Rev ARM from the NMR structure was aligned with the ARM of one Rev monomer in the crystal structure (backbone r.m.s.d. = 1.0 Å), unambiguously placing the RNA relative to the Rev dimer. Four basepairs of A-form RNA with idealized geometry were added to the end of the IIB element to extend the RNA from the 34 nt in the NMR structure to the 42 nt RNA that cooperatively binds the Rev dimer (Fig. 1b).

To model RNA bound to the Rev hexamer, we began by docking a 42 nt RNA to one Rev dimer in the same way as above. Previous work with model RNAs suggests that further extending the stem IIB RNA (to 64 nt) facilitates binding of more than one Rev dimer<sup>18</sup>. We therefore extended the modeled 42 nt an additional 9 base pairs of A-form helix to contact the second Rev dimer in the hexamer. Additional A-form RNA would not contact the third Rev dimer, consistent with the many stem-loops and stem-junctions in the RRE, so the third dimer was modeled to bind RNA similar to the first dimer. We expect that the complex folding of the RRE will adopt a different structure than that shown, but that the proximity and orientation of binding sites will be retained.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

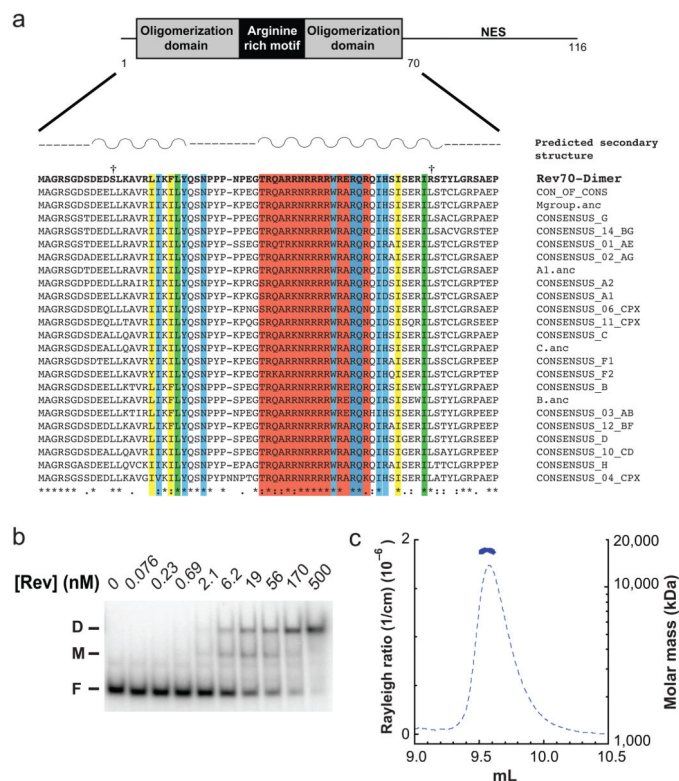
We are grateful for assistance with crystallography and structure determination from Zach Newby, Pascal Egea, Janet Finer-Moore, Robert Stroud, Miles Pufall, JJ Miranda, Dong Young Kim, Brett Kaiser, Melanie Brewer, Lesa Beamer, James Holton, and George Meigs. We also thank David Booth, Bhargavi Jayaraman, John Gross, Yifan Cheng, Raul Andino, Geeta Narlikar and Harmit Malik for helpful discussions and critical reading of the manuscript. M.D.D. was supported by a Howard Hughes Medical Institute predoctoral fellowship.

## References

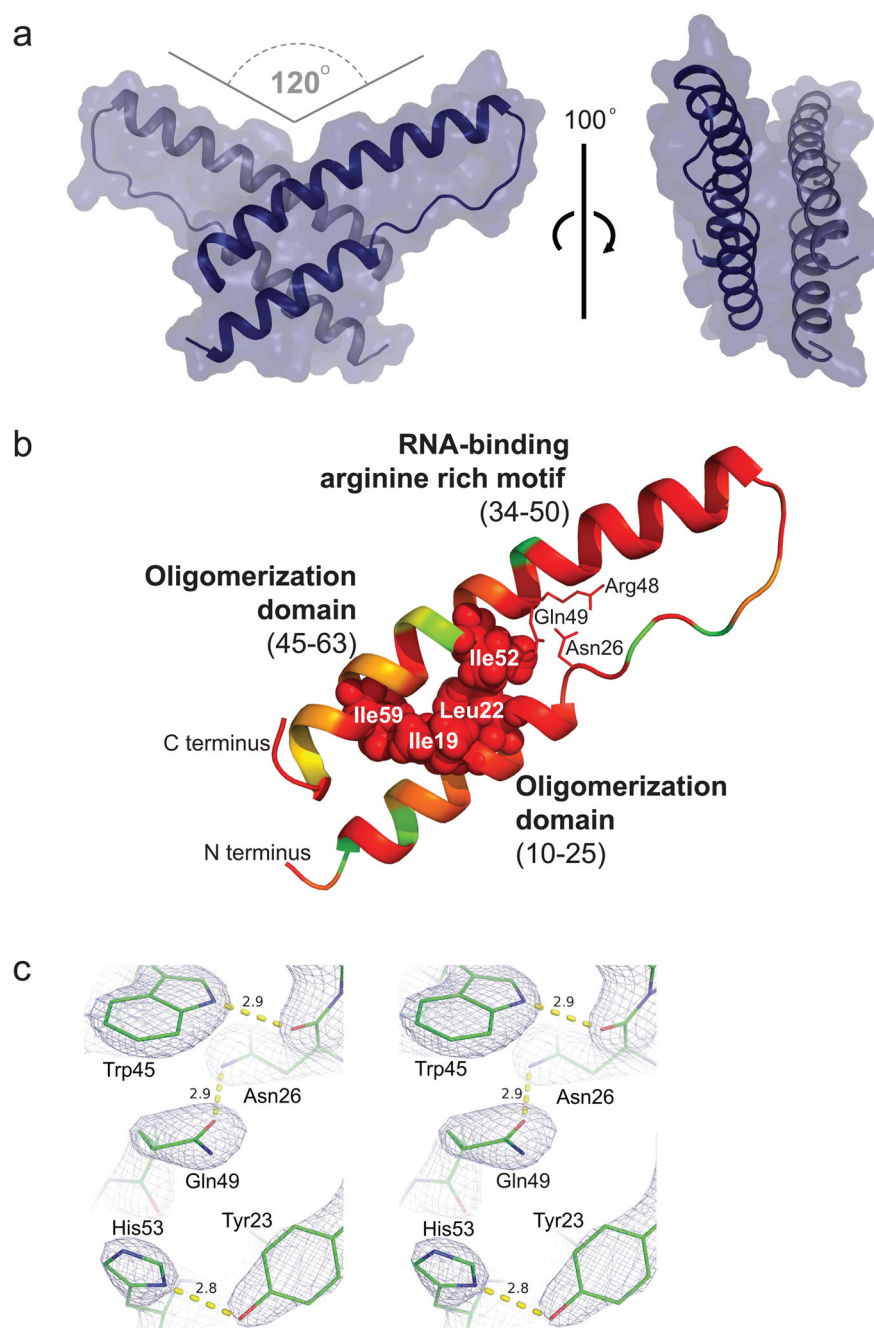
1. Cullen BR. Nuclear mRNA export: insights from virology. *Trends Biochem Sci.* 2003; 28:419–24. [PubMed: 12932730]
2. Pollard VW, Malim MH. The HIV-1 Rev Protein. *Annu Rev Microbiol.* 1998; 52:491–532. [PubMed: 9891806]
3. Fornerod M, Ohno M, Yoshida M, Mattaj IW. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell.* 1997; 90:1051–1060. [PubMed: 9323133]

4. Malim MH, Cullen BR. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell*. 1991; 65:241–48. [PubMed: 2015625]
5. Mann DA, et al. A molecular rheostat: Co-operative rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J Mol Biol*. 1994; 241:193–207. [PubMed: 8057359]
6. Daugherty MD, D’Orso I, Frankel AD. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell*. 2008; 31:824–34. [PubMed: 18922466]
7. Daugherty MD, Booth DS, Jayaraman B, Cheng Y, Frankel AD. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proc Natl Acad Sci USA*. 2010; 107:12481–12486. [PubMed: 20616058]
8. Cook KS, et al. Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res*. 1991; 19:1577–83. [PubMed: 2027765]
9. Heaphy S, et al. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell*. 1990; 60:685–93. [PubMed: 1689218]
10. Malim MH, et al. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell*. 1990; 60:675–83. [PubMed: 2406030]
11. Kjems J, Brown M, Chang DD, Sharp PA. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc Natl Acad Sci USA*. 1991; 88:683–7. [PubMed: 1992459]
12. Heaphy S, Finch JT, Gait MJ, Karn J, Singh M. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich “bubble” located within the rev-responsive region of viral mRNAs. *Proc Natl Acad Sci USA*. 1991; 88:7366–70. [PubMed: 1871141]
13. Pond SJ, Ridgeway WK, Robertson R, Wang J, Millar DP. HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc Natl Acad Sci USA*. 2009; 106:1404–8. [PubMed: 19164515]
14. Battiste JL, et al. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science*. 1996; 273:1547–1551. [PubMed: 8703216]
15. Xu W, Ellington AD. Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc Natl Acad Sci USA*. 1996; 93:7475–80. [PubMed: 8755498]
16. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA*. 2005; 11:1848–57. [PubMed: 16314457]
17. Landt SG, Ramirez A, Daugherty MD, Frankel AD. A simple motif for protein recognition in DNA secondary structures. *J Mol Biol*. 2005; 351:982–94. [PubMed: 16055152]
18. Zimmel RW, Kelley AC, Karn J, Butler PJ. Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. *J Mol Biol*. 1996; 258:763–77. [PubMed: 8637008]
19. Jain C, Belasco JG. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol Cell*. 2001; 7:603–614. [PubMed: 11463385]
20. Dimattia MA, et al. Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proc Natl Acad Sci USA*. 2010; 107:5810–4. [PubMed: 20231488]
21. Auer M, et al. Helix-loop-helix motif in HIV-1 Rev. *Biochemistry*. 1994; 33:2988–96. [PubMed: 7510518]
22. Edgcomb SP, et al. Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes. *Protein Sci*. 2008
23. Vallone B, Miele AE, Vecchini P, Chiancone E, Brunori M. Free energy of burying hydrophobic residues in the interface between protein subunits. *Proc Natl Acad Sci USA*. 1998; 95:6103–7. [PubMed: 9600924]
24. Williamson JR. Cooperativity in macromolecular assembly. *Nat Chem Biol*. 2008; 4:458–65. [PubMed: 18641626]

25. Khorasanizadeh S, Rastinejad F. Nuclear-receptor interactions on DNA-response elements. *Trends Biochem Sci.* 2001; 26:384–90. [PubMed: 11406412]
26. Hong M, et al. Structural basis for dimerization in DNA recognition by Gal4. *Structure.* 2008; 16:1019–26. [PubMed: 18611375]
27. Chen L, Glover JN, Hogan PG, Rao A, Harrison SC. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature.* 1998; 392:42–8. [PubMed: 9510247]
28. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol.* 1980; 136:225–70. [PubMed: 7373651]
29. Larkin MA, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–8. [PubMed: 17846036]
30. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999; 292:195–202. [PubMed: 10493868]
31. Monecke T, et al. Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science.* 2009; 324:1087–91. [PubMed: 19389996]
32. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* 1997; 276:307–326.
33. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr.* 1999; 55:849–61. [PubMed: 10089316]
34. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1948–54. [PubMed: 12393927]
35. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997; 53:240–55. [PubMed: 15299926]
36. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–32. [PubMed: 15572765]
37. McCoy AJ, et al. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
38. Davis IW, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007; 35:W375–383. [PubMed: 17452350]
39. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994; 50:760–3. [PubMed: 15299374]
40. Delano, WL. The PyMOL molecular graphics system. Vol. 0.99. Delano Scientific; 2006.

**Figure 1.**

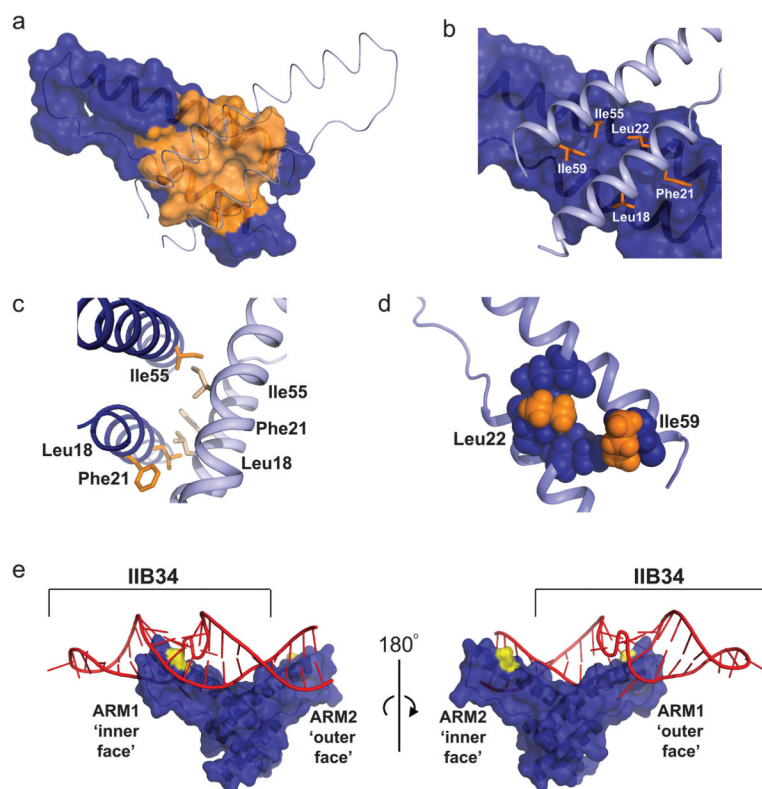
Rev domain structure and protein dimerization. **(a)** Domain structure of the Rev protein, including the disordered C-terminal region containing the nuclear export sequence (NES). Below is a consensus alignment of the first 70 residues of Rev from reference subtypes (Los Alamos HIV sequence database (<http://www.hiv.lanl.gov/>)) represented using the ClustalW program<sup>29</sup>. The top sequence is the one used for crystallization, with oligomerization mutants L12S and L60R indicated by daggers, and above is the predicted secondary structure<sup>30</sup>. Regions predicted to be helical are indicated by wavy lines and coiled regions by dashed lines. Highlighted are residues shown in the Rev dimer structure to be important for Rev monomer stability and folding (blue), Rev dimer formation (yellow), or both (green)<sup>19</sup>. Residues in red are the highly conserved residues of the ARM. **(b)** Representative gel shift assay data using purified Rev70-Dimer at the concentrations indicated and radiolabeled IIB42 RNA. The species indicated are: F, Free RNA; M, Rev monomer; D, Rev dimer. The dissociation constant ( $K_D = 14 \pm 6.7$  nM) and Hill coefficient ( $\eta = 1.1 \pm 0.027$ ) were calculated using the equation:  $FractionBound = [Rev]^\eta / (K_D^\eta + [Rev]^\eta)$ . Data are presented as the mean  $\pm$  s.d. from three replicates. **(c)** Measured multi-angle light scattering (MALS) (dashed lines, left axis) and calculated molar masses (solid line, right axis) determined from the major peak from size exclusion chromatography (SEC) of 250  $\mu$ M Rev70-Dimer protein in crystallization buffer containing 0.5 M  $SO_4$ . The measured mass of 17 kDa corresponds to two Rev monomers (8.5 kDa each).

**Figure 2.**

Overall structure of the Rev dimer and monomer. **(a)** Two views of a surface representation of the Rev dimer, with the two monomers shown as ribbons (dark and light blue) and the crossing angle indicated. **(b)** Each of the four Rev monomers has a folded core from residues 9–63 as shown, with structural and functional regions and amino acid numberings indicated. All monomer structures are highly similar, with a root mean squared deviation (r.m.s.d.) of 0.5–1 Å for all pairwise alignments of backbone atoms from residues 9–63, and a backbone r.m.s.d of 1.0 Å when aligned to the recent Fab-bound Rev structure<sup>20</sup>. Amino acid conservation among 1201 HIV-1 isolates in the Los Alamos HIV sequence database is

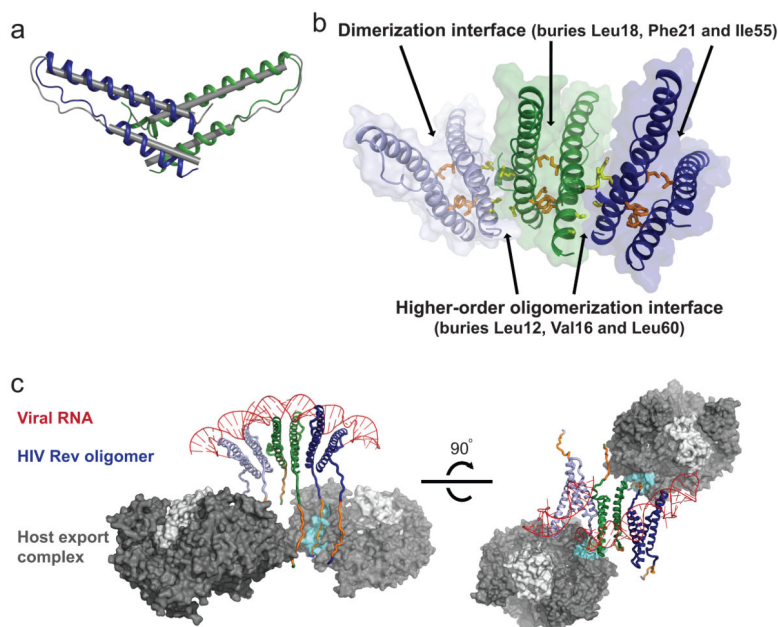
indicated on the ribbon diagram, ranging from lowest conservation (26%, green) to highest (100%, red). Hydrophobic and polar residues that stabilize the monomer structure are shown as spheres and sticks and are colored by sequence conservation. **(c)** Stereo view of a  $\sigma_A$ -weighted  $2F_{\text{obs}}-F_{\text{calc}}$  map contoured at  $2.0\sigma$  of the hydrogen bonding network stabilizing the Rev monomer structure. Distances between hydrogen-bonding donor and acceptor atoms are indicated ( $\text{\AA}$ ).



**Figure 3.**

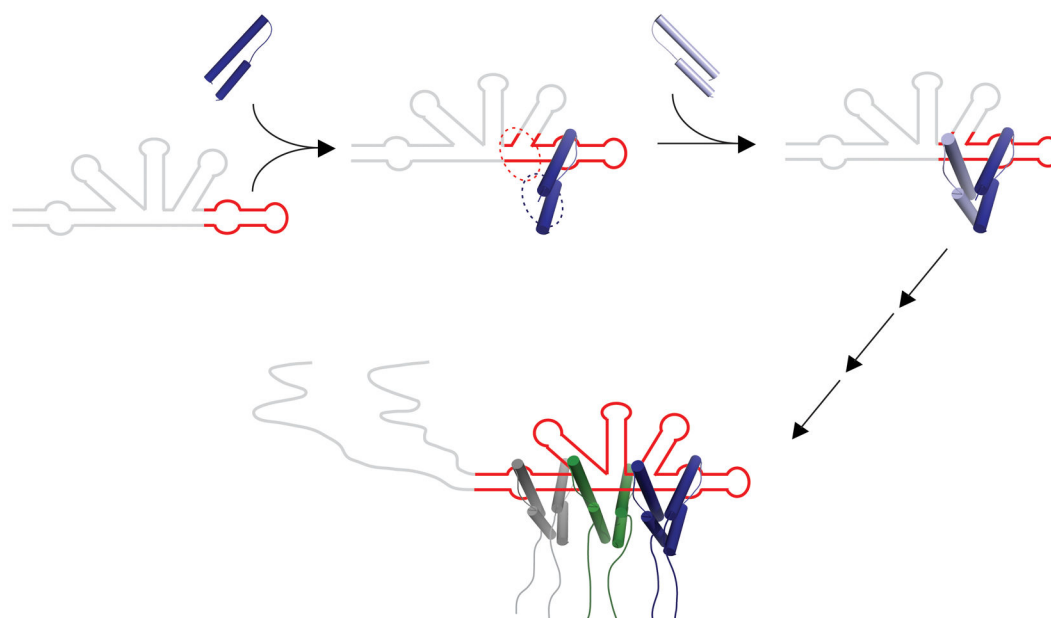
The Rev dimer interface mediates cooperative RNA recognition. **(a)** Cutaway of the dimer interface showing the surface of one monomer (blue), and the predicted solvent accessible surface area buried upon dimerization (orange) with the second monomer (light blue ribbon). **(b)** The five critical hydrophobic residues that comprise the dimer interface, shown as sticks in one monomer projecting onto the surface of the second. **(c)** Three hydrophobic residues mediate symmetric interactions at the dimerization interface. Individual monomers are colored in light and dark blue. **(d)** Residues that form the core of the monomer structure (Leu22 and Ile55) are buried further upon dimerization. Nonpolar surface area of these residues is partially buried in the monomer (blue) and additionally buried in the dimer (orange). **(e)** Two views of the Rev dimer (blue) modeled with the 34 nt stem IIB (labeled IIB34; red) from the NMR structure of an ARM peptide–RNA complex<sup>14</sup>. The peptide helix was aligned with ARM1 (r.m.s.d = 1.0 Å), and four additional base pairs of idealized A-form RNA were appended to the end of the RNA (red). The 42 nt RNA that binds the Rev dimer cooperatively (Fig. 1) contains bulged nucleotides that are missing from this model, but the trajectory and length of the helix should be similar. The positions of Asn40 are shown for each monomer (yellow), showing how the ‘inner face’ of ARM1 positions Asn40 to contact stem IIB while the ‘outer face’ of ARM2 (without Asn40) is positioned to contact the adjacent RNA site.





**Figure 4.**

Arrangement of the Rev oligomer and model of the Rev–RRE RNP and Crm1 interaction. **(a)** Comparison of the higher-order oligomerization interface seen between asymmetric units in the crystal (blue and green cartoon) and a recent structure of Rev with a monoclonal Fab bound to the dimer interface<sup>20</sup> (grey cylinders) (backbone r.m.s.d = 1.9 Å). **(b)** A Rev hexamer generated using the arrangement of three Rev dimers in the crystal (blue, green and light blue cartoons). **(c)** Two views of a jellyfish model of the Rev hexamer bound to RNA and its interaction with Crm1. RNA was docked onto one dimer of the Rev hexamer as in Fig. 2e, and extended an additional 9 base pairs of A-form helix to contact the second dimer (seen in the right panel, green dimer) based on previous studies with model RNAs<sup>18</sup>. An additional fragment of RNA was docked in the same manner as in Fig. 2e to the third dimer (light blue) to model complete binding of the Rev hexamer to the RRE (see Methods). The disordered C-termini containing the NES (residues 73–83; orange) were modeled as random coil extensions and are seen to project away from the RNA-binding ARMs. The NES sequences were fit into the NES binding site (cyan) of Crm1 based on the structure of a Crm1–RanGTP complex (dark and light grey surfaces, respectively) with Snurportin1 bound at the NES site<sup>31</sup>. Two Crm1 complexes are shown for scale and suggest that no more than two are likely to be accommodated in this arrangement.



**Figure 5.**

Model for oligomerization-mediated cooperative assembly. Binding of Rev (top left, blue cartoon) to the RRE (grey) is initiated at the stem IIB site<sup>13</sup> (red). The resulting complex (top middle) exposes the dimerization interface (blue dashed oval) for binding by a second Rev monomer. Only when a second Rev-binding RNA site (red dashed oval) is properly oriented relative to the stem IIB site will the second Rev monomer be able to cooperatively assemble into the RNP using this composite protein–protein and protein–RNA interaction surface. Further oligomerization and protein–RNA interactions will likewise facilitate cooperative assembly of the complete hexameric Rev–RRE RNP (bottom). At this point, the complex is organized for nuclear export, with the NESs of the Rev subunits projecting downward (as in Fig. 4c) for interaction with Crm1 and the viral RNA bound to opposite face of the Rev oligomer.

**Table 1**

Crystallographic data collection and refinement statistics.

	Native	SeO <sub>4</sub> -MAD		Br-SAD
Data collection				
Space group	P6 <sub>4</sub> 22	P6 <sub>4</sub> 22		P6 <sub>4</sub> 22
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	115.8, 115.8, 81.2	115.9, 115.9, 81.4		115.9, 115.9, 81.1
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0		90.0, 90.0, 120.0
		<i>Peak</i>	<i>Remote</i>	<i>Peak</i>
Wavelength	1.1159	0.9792	0.9715	0.9202
Resolution (Å)	50.0–2.5	50.0–2.8	50.0–2.8	50.0–3.2
<i>R</i> <sub>merge</sub>	5.1 (50.7)	6.0 (36.7)	5.7 (33.7)	13.2 (56.6)
<i>I</i> / <i>σI</i>	35.2 (3.4)	29.9 (3.9)	32.5 (4.8)	14.9 (3.3)
Completeness (%)	99.7 (99.0)	100 (99.9)	100 (99.9)	100 (100)
Redundancy	7.6 (6.1)	14.3 (11.1)	14.8 (12.2)	12.2 (12.8)
Refinement				
Resolution (Å)	47.1–2.50			
No. reflections	11031			
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.226/0.261			
No. atoms				
Protein	1868			
Ligands/ions	37			
Water	59			
<i>B</i> -factors				
Protein	62.7			
Ligand/ion	84.0			
Water	54.1			
R.m.s deviations				
Bond lengths (Å)	0.004			
Bond angles (°)	0.677			

A single crystal was used to collect each of the native, multi-wavelength anomalous diffraction (MAD) and single-wavelength anomalous diffraction (SAD) datasets. Values in parentheses are for the highest-resolution shell.